Methods for Building High Quality and Comprehensive MS/MS Libraries

Xiaoyu Yang, Pedatsur Neta, Yamil Simón-Manso, Yuxue Liang, Lisa Kilpatrick, Jeri Roth, Maria Lorna A. De Leoz, Wei Mi, Stephen E. Stein

National Institute of Standards and Technology, Gaithersburg, MD

➢ Overview:

- Purpose: Develop high quality and comprehensive MS/MS libraries
- Methods: Consensus spectra were generated from experimental data by a clustering algorithm; Spectral quality was ascertained based on fragment annotation, mass accuracy, base peak intensity, and S/N ratio etc.
- Results: Libraries of 7,020 compounds, 15,517 precursor ions and 123,781 spectra have been built.

➢ Introduction:

- Mass spectral library searching is an effective method for chemical identification because of its reliable searching results and fast searching speed;
- The quality, size, and completeness of a library are the key to successful searching.
- We aim to provide the largest possible collection of high quality reference spectra of biologically and environmentally relevant compounds such as metabolites, bioactive peptides, lipids, sugars, glycans, pesticides, surfactants, and various contaminants.

➢ Methods:

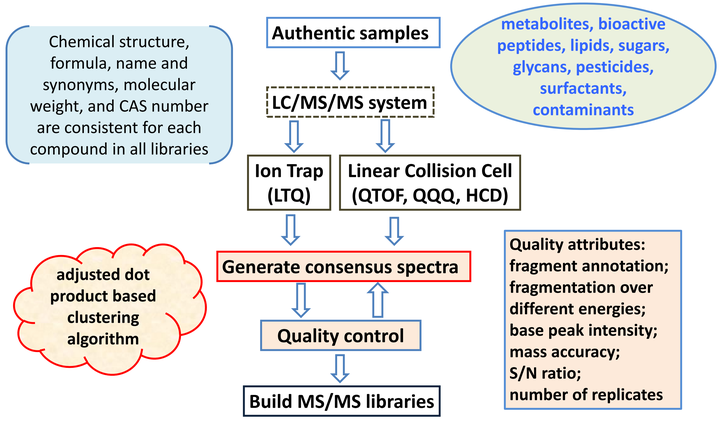

Fig. 1. Procedure of building MS/MS library

➢ Results:

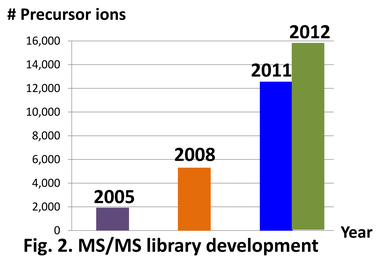

Fig. 2. MS/MS library development

7,020 Compounds
15,517 Precursor Ions
123,781 Spectra
90% Positive Ion Spectra
10% Negative Ion Spectra

Instrument Type	Precursor Ions
Ion Trap	12,192
Collision Cell (QTOF, QQQ, HCD)	9,432

Major precursor types:
$[M+H]^+$, $[M+2H]^{2+}$, [M-H]-, $[M+Na]^+$, $[M+NH4]^+$, $[Cat]^+$, [An]-, [p-H2O], [p-NH3]
Peptide charge states: 1 to 10 and -1 to -4

❖ Quality Control #1: making consensus spectra

- An adjusted dot product based clustering algorithm was used to group similar spectra into the same cluster and created one consensus spectrum from each cluster; The best consensus spectrum was picked for the library.
- Noise peaks were removed using a voting algorithm.
- Example: two clusters were generated for Palatinose [M+NH4]+ (Fig. 3); The spectrum in Fig. 3A was kept in the library.

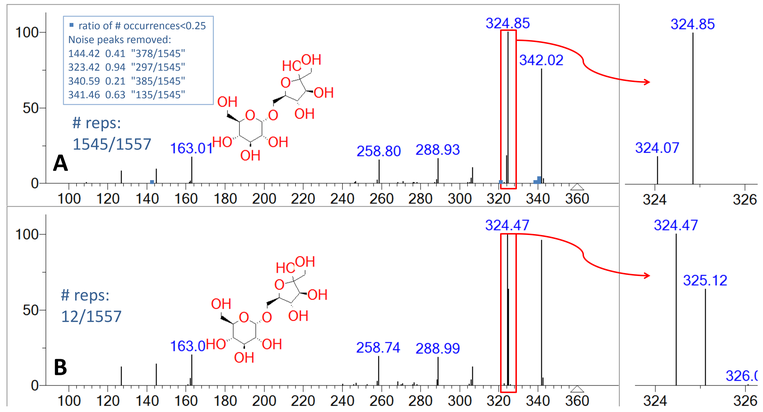

Fig. 3. Generating consensus spectra for Palatinose [M+NH4]+ on LTQ

❖ Quality Control #2: peak annotation

- Each spectrum was ascertained that all major peaks are assigned to acceptable fragmentation product ions from the known precursor structure or peptide sequence (Fig. 4).

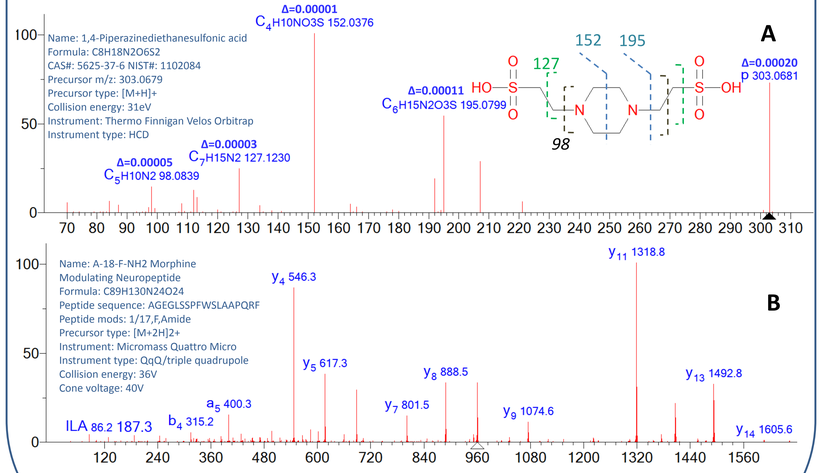

Fig. 4. Annotating spectrum peaks based on precursor structure (A) or peptide sequence (B)

❖ Quality Control #3: fragmentations over different energies

- Each spectrum was manually inspected to ascertain that peak intensities vary with collision energy in a reasonable progression.

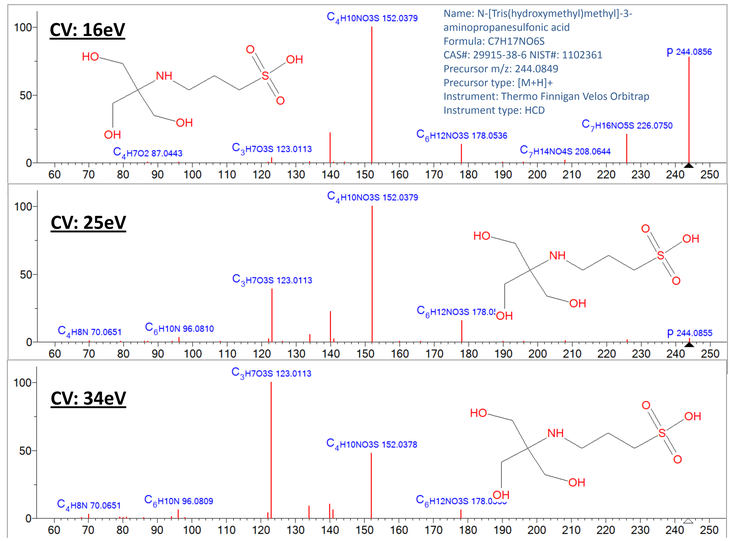

Fig. 5. Peak intensity changes with collision energy

❖ Quality Control #4: different instrument, different precursor types, etc.

- The same sample was run on the different instruments (Fig. 6) at positive or negative modes with various precursor types.

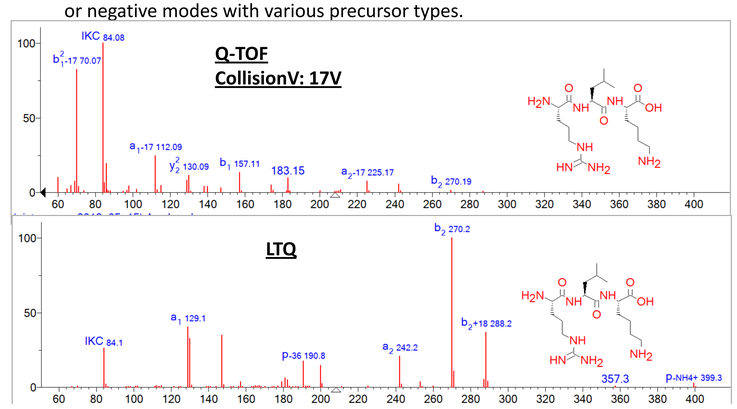

Fig. 6. A tripeptide RLK/2 was analyzed on LTQ and QTOF instruments

➢ **Conclusions:** High quality and comprehensive libraries are being developed for metabolites, peptides, lipids, sugars, glycans, pesticides, surfactants, and contaminants, etc.